Postal Regulatory Commission
Submitted 2/19/2021 3:45:25 PM
Filing ID: 116120
Accepted 2/19/2021

BEFORE THE
POSTAL REGULATORY COMMISSION
WASHINGTON, D.C. 20268–0001

PERIODIC REPORTING (PROPOSAL SIX)	Docket No. RM2020-13

RESPONSES OF THE UNITED STATES POSTAL SERVICE TO QUESTIONS 1-7 OF CHAIRMAN'S INFORMATION REQUEST NO. 6
(February 19, 2021)

The United States Postal Service hereby provides its responses to the above listed questions of Chairman's Information Request No. 6, issued February 10, 2021. The questions are stated verbatim and followed by the response.

Respectfully submitted,

UNITED STATES POSTAL SERVICE

By its attorney:

Eric P. Koetting

475 L'Enfant Plaza West, S.W.
Washington, D.C. 20260-1137
(202) 277-6333
eric.p.koetting@usps.gov
February 19, 2021

RESPONSE OF THE UNITED STATES POSTAL SERVICE TO CHAIRMAN'S INFORMATION REQUEST NO. 6

1. Please refer to the Reply Variability Report providing that "[e]ven at the 1% tails, it is perhaps notable that none of the values in Table 6 are clearly erroneous, particularly for [delivery bar code sorter (DBCS)] operations. Observations for [Automated Flats Sorting Machine (AFSM)] 100 operations less than half the median or [Flats Sequencing System (FSS)] observations more than twice the median may both be regarded as at least being anomalous[.]" Reply Variability Report at 15. Please also refer to Response to CHIR No. 1 that states "setting operation-specific productivity cutoffs based on machine characteristics was rejected, as it is not possible to set unambiguous cutoffs based on available information on machine throughput and staffing levels, particularly for AFSM 100 and FSS equipment subject to variable throughput and staffing levels[.]" Response to CHIR No. 1, question 6.b.

a. Please confirm that that for DBCS, AFSM 100, or FSS equipment (operations), setting operation-specific productivity cutoff values is rejected. If not confirmed (or partially confirmed), please discuss the criteria for determining operation-specific productivity cutoffs and provide their specific values for all or any of the three referenced types of machine operations.

b. If question 1.a. is confirmed for flats (AFSM 100 and FSS) operations, please discuss the reasoning underlying the conclusion that the observations "for AFSM 100 operations less than half the median or FSS observations more than twice the median may both be regarded as at least being anomalous." Reply Variability Report at 15.

RESPONSE:

a. Partly confirmed. The Proposal Six productivity cutoffs shown in Table 1 of the Variability Report (at page 21) are operation-specific, but based on statistical criteria rather than machine operating characteristics such as nominal throughput or staffing levels, as stated in the cited response to ChIR No. 1, question 6b.

b. While factors such as machine characteristics and staffing levels cannot set *unambiguous* cutoffs, as previously stated, they are informative for identifying observations that are "at least… anomalous." Extremely high productivities may be technically infeasible to the extent they imply clearly insufficient staffing levels to operate machines (e.g., less than one workhour per hour of runtime for machines operated by multiple employees). Low productivities may not be technically infeasible as such, as there is not a strict lower bound on technical capabilities, but they can imply unreasonable amounts of idle labor notwithstanding short-term labor inflexibilities faced by Postal Service mangers.

Less extreme productivity values may be technically feasible at some level, but still anomalous. Typical productivities should be materially lower than a productivity ceiling such as a maximum machine throughput divided by a staffing level, since a substantial number of workhours are incurred in overhead and other activities not involving handling of mail. This is the basis for judging certain productivities in the outer tails of the productivity distributions to be anomalous (if not erroneous). In practice, the productivity ceiling will be unattainable due to overhead hours in the operations. (On a systemwide basis, overhead activities are 28 percent of DBCS labor, 30 percent of AFSM 100 labor, and 35 percent of FSS labor; see USPS-FY20-7 part 7.xlsx, worksheet "VII-3a.ovhead factors-MODS1&2.) Actual attainable productivities will depend on actual staffing levels and overheads, which may vary by site from system averages and/or operational minimums, and are not directly observable (or measurable with sufficient statistical precision at the site level).

For example, an observed AFSM 100 productivity of 4,000 TPF per workhour would be technically infeasible in that it exceeds a throughput-staffing ceiling. Such an observation would be eliminated from the regression samples at any of the screening levels. Nominal throughput of an AFSM 100 machine is approximately 17,000 pieces per hour of operation (Docket No. R2006-1, USPS-T-42); full complement for an AFSM 100 machine (including feeders, sweepers, and employees working prep operations) is seven or more—e.g., an AFSM 100 AI machine would be fully staffed with a feeder, one or two sweepers, and five employees working the prep subsystem. An observed AFSM 100 productivity of 2,312 TPF per workhour — the AFSM 100 upper 1% cutoff — may be technically feasible (i.e., as it is less than 17,000 divided by seven) but would imply that the site incurred unusually low overhead hours (possibly among other anomalies). Thus, an AFSM productivity at or above the 1% cutoff may be reasonably judged to be anomalous.

2. Please refer to Table 6 of the Reply Variability Report that presents the analysis of the productivity screen cutoffs (by machine operation) for the 1%, 5%, and 10% cutoff values. *Id.* Please also refer to the Response to CHIR No. 5 presenting "effects of [two] alternative screens on estimated variabilities" in Table 4. Response to CHIR No. 5, question 7.c.iv.

a. Please provide the results of the productivity screen analyses, (similar to what was provided in Table 6 of the Reply Variability Report) for the two alternative productivity cutoffs suggested in Response to CHIR No. 5, question 7.c.iv., Table 4 (first, 5% tails, FY 2016-2019 data and, second, 5% tails, computed by month, FY 2016-2019 data).

b. For the results of productivity screen analysis provided in question 2.a., please indicate which productivity values could be seen as at least being anomalous and explain why.

RESPONSE:

a. Please see the workbook CHIR.6.Q2a.xlsx, electronically attached to this response.

b. The productivity percentiles based on FY2016-2019 data in CHIR.6.Q2a.xlsx show similar spreads of the outer tail values relative to the medians as the analysis in the Reply Variability Report. The cutoffs themselves shifted down overall, with the exception of some peak-month values for flat operations, though the 1% tail values for AFSM 100 and FSS operations are at levels where they may be seen as anomalous as described in the response to Question 1(b) of this Information Request. In addition, the downward shift of the productivity distribution is associated with declining volumes that may limit the extent to which sites can operate equipment at nominal throughput, which would tend to reduce the effective productivity ceilings.

RESPONSE OF THE UNITED STATES POSTAL SERVICE TO CHAIRMAN'S INFORMATION REQUEST NO. 6

3. Please refer to the Variability Report that states: "We found that the estimated elasticities for workhours were somewhat sensitive to the inclusion of outliers with unusual values for labor productivity in regressions using unscreened data." Variability Report at 21. Please also refer to Table 1 below.

a. Please confirm that Proposal Six flats variabilities, estimated from FSS and AFSM 100 workhour regression equations (with lag, seasonal variables, and for the FY 2016-2019 sample period), generally increase as the productivity screen cutoffs become more restrictive. *See* Table 1 below. If confirmed, please explain why the removal of a higher number of observations generally results in higher variabilities.

b. If question 3.a. is not confirmed, please discuss the relationship, if any, between productivity screen cutoffs and the magnitude of the variability estimates. If there is no such relationship, please explain why higher estimated variabilities are generally associated with more restrictive productivity screen cutoffs. For example, see numbers in columns (7) and (8) of Table 1 below, which are always positive for flats.

c. Please discuss why variabilities estimated from the AFSM 100 workhour regression equations appear to be more "sensitive" to changes in productivity screen cutoffs than variabilities for DBCS and FSS operations.[1]

Table 1. Workhour Variability Estimates Sensitivity

	No screen		1% screen		5% screen (Proposal Six)		Sensitivity	
	coeff	se	coeff	se	coeff	se	Coefficient Diff.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7) =(5)-(1)	(8) =(5)-(3)
DBCS	0.821	0.059	0.992	0.035	0.976	0.032	15.6%	-1.5%
AFSM 100	0.518	0.144	0.722	0.095	0.774	0.091	25.5%	5.2%
FSS	0.732	0.110	0.750	0.092	0.804	0.070	7.2%	5.4%

Source: Library Reference USPS-RM2020-13/1, September 15, 2020, folder "Analysis," Excel file "results_seasonal.xlsx."

[1] As illustrated in Table 1, the difference between the 5% screen and no screen estimates for AFSM 100 operations is 25.5 percent.

RESPONSE OF THE UNITED STATES POSTAL SERVICE TO CHAIRMAN'S INFORMATION REQUEST NO. 6

RESPONSE:

a. Confirmed that the point estimates of the AFSM 100 and FSS elasticities are higher in the runs with stricter screens. However, a number of the differences shown in Table 1 are statistically insignificant — e.g., the differences between the 1% and 5% screens for AFSM 100 and FSS, and the difference between no screen and the 1% screen for AFSM 100, are less than one estimated standard error — and the DBCS elasticity is non-increasing between the 1% and 5% screens. Thus, the increase in the point estimates is not a result that necessarily can be generalized. The largest and most qualitatively significant effects are from screening extreme outliers for AFSM 100 and DBCS operations in the step from no screen to the 1% screen. See also the response to part (c), below.

b. Not applicable.

c. The Commission's table shows the effects of two steps of screening levels with distinct effects on the estimated variabilities. In the first step, between the no-screen and 1% screen alternatives, both the AFSM 100 and DBCS elasticities exhibit greater sensitivity to imposing the 1% screen than the FSS elasticity. The differential effect appears to result from the presence of more extreme productivity outliers in the AFSM 100 and DBCS samples prior to screening. See USPS-RM2020-13-1, file "analysis_seasonal.txt" summary statistics of the productivity distribution including extreme high and low values.

In the second step, increasing the stringency of the productivity screen from the 1% to the 5% tails of the productivity distribution, the effect on the AFSM 100 elasticity is qualitatively and quantitatively similar to that on the FSS elasticity. Insofar as the DBCS elasticity estimate is lower in the 5% screen case than the 1% screen case, it does not appear to be the case that more restrictive cutoffs necessarily increase the elasticity estimates. For all three groups, the differences in the elasticities for the 1% and 5% screens are less than one estimated standard error.

RESPONSE OF THE UNITED STATES POSTAL SERVICE TO
CHAIRMAN'S INFORMATION REQUEST NO. 6

4. Please refer to Response to CHIR No. 4 that states "[t]he main reason the Postal Service did not examine models with the full set of lags is due to the likelihood that such a specification would encounter multicollinearity issues leading to statistically unreliable estimates of the coefficients on many or most of the lagged TPF variables. Additionally, there is little theoretical or operational basis for including the second through eleventh lags, compared to the first and twelfth lags." Response to CHIR No. 4, question 3.d. Please also refer to the Variability Report that provides: "[t]est statistics for the joint inclusion of the lagged [total pieces fed (TPF)] and monthly dummy variables strongly reject the null hypothesis that lagged and seasonal effects [included into the extended equations] are jointly zero . . . [although for] AFSM 100, the joint test that the lagged TPF coefficients are zero does not reject the null hypothesis at standard significance levels (p-value 0.13)." *Id.* at 23-24, 24 n.10.

a. Please discuss in details [sic] why "there is little theoretical or operational basis for including the second through eleventh lags [into Proposal Six's econometric model], compared to the first and twelfth lags." Response to CHIR No. 4, question 3.d.

b. Please discuss the reasons for estimating the AFSM 1000 [sic] variabilities from the extended regression equation with the first and the twelfth lags of monthly TPF (and not with any other lags) although the provided joint significance test did "not reject the null hypothesis at standard significance levels (p-value 0.13)." Variability Report at 24 n.10.

c. Please provide academic references (the Postal Service relied on when specifying the regression workhour models) that discuss how to determine the appropriate number of lags of an independent variable in a dynamic regression model.

RESPONSE:

a. The basis for the inclusion of the first and twelfth lags is discussed in the Variability Report at page 20, cited in the response to ChIR No. 4, question 3(d), and in the Reply Variability Report at 9-10. These are intended to capture potential effects of short-term staffing constraints as well as the use of SPLY data in operational management. The staffing constraints would tend to be less binding over longer intervals as scheduling can adapt to near-term volume expectations. SPLY information will tend to reflect seasonality of operations, where other lags that may reflect different seasonal peaks or troughs would not. Finally, a data consideration is that while multicollinearity problems do not arise with the inclusion of two lags of TPF in the Proposal Six specification, as noted in the response to ChIR No. 2, question 4, the lagged (log) TPF variables are highly correlated and addition of additional lags may introduce multicollinearity problems. The multicollinearity issue is noted in the Baltagi text cited in the responses to ChIR No. 2, Question 2 and ChIR No. 4, question 3. Badi H. Baltagi, *Econometrics* (Springer-Verlag, 2008) at 129.

b. Both lags were retained in the AFSM 100 variabilities because the null hypothesis that the coefficient on the first lag of TPF is zero was rejected at standard significance levels (p-value less than 0.01), and the null hypothesis for the joint test could be rejected in other sample periods (e.g., FY2015-2019) at standard significance levels (p-value less than 0.05). Including the statistically insignificant lag does not lead to statistical bias or inconsistency, as it does not impose a restriction relative to a model excluding one or both lags, and would minimize the need to pretest the AFSM 100 model specification during the course of updating the models to incorporate new data.

c. As is noted in the Baltagi text (*id.*), lagged independent variables are additional regressors satisfying "classical assumptions." Badi H. Baltagi, *Econometrics* (Springer-Verlag, 2008) at 129. Testing distributed lag specifications is thus a matter of testing restrictions on regression coefficients. See., e.g., *id.* at 130-134. For Proposal Six, inclusion of the first and twelfth lags was tested against the alternative of no lags.

RESPONSE OF THE UNITED STATES POSTAL SERVICE TO CHAIRMAN'S INFORMATION REQUEST NO. 6

5. Please refer to Docket No. R97-1, Appendices to Opinion and Recommended Decision, Volume 2, May 11, 1998, Appendix F (Docket No. R97-1 Opinion)[2] that states "the estimator for the fixed effects . . . shows that the fixed effects [dummies] will include all of the difference between the average labor processing times for the facilities that is not captured by differences in the averages for piece handlings and the controls. There is nothing about the estimator for the fixed effects that prevents them from reflecting volume-variable indirect effects at the facility level." Docket No. R97-1 Opinion, Appendix F at 42.

a. Please confirm that the facility-specific fixed effects dummy variables could contain volume-variable indirect effects of TPF on workhours at the facility level.

b. If question 5.a. is confirmed, please explain whether it is appropriate to assume that such volume-variable indirect effects are not captured (or not fully captured) in the variability estimates in Proposal Six.

c. If question 5.a. is not confirmed, please explain what prevents the fixed effects dummy variables from reflecting volume-variable indirect effects of TPF on workhours at the facility level.

RESPONSE:

a. Not confirmed.

b. Not applicable.

c. A correct interpretation of the fixed effects (or estimated site-specific intercepts) is that they capture the effects of variables that are constant over time (but not over sites) that are *excluded* from the regression specification. Consider the model:

$$y_{it} = \alpha + \beta x_{it} + \gamma z_i + u_{it},$$

where x_{it} is a vector of observed variables that vary over time t and sites i, z_i is a vector of unobserved variables that vary over sites but not time, u_{it} is a random

---

[2] This document is available in the Postal Rate Commission Archives (1971-2004) at *https://www.prc.gov/prcarchive/viewpdf.aspx?docid=26815*.

disturbance term, and α, β, and γ are parameters. Without observing z_i, it is not possible to estimate γ. However, it is possible to estimate the combination δ_i=$\alpha + \gamma z_i$. Substituting leads to the variable-intercept model:

$$y_{it} = \delta_i + \beta x_{it} + u_{it}.$$

Thus, site dummy variables in a "one-way" fixed-effects model estimate the effects of site-specific omitted variables plus the overall intercept (noting that the overall intercept α is not separately identifiable). Similarly, time dummy variables would estimate the effects of omitted time-varying variables. (The Proposal Six models do not employ time dummy variables; see the response to ChIR No. 3, question 3(b) for related discussion.) This is the basis for the fixed-effects model's well-known application for mitigating omitted-variables bias. See, e.g., Cheng Hsiao, *Analysis of Panel Data* (Cambridge University Press, 1986) at 25-27. Effects of the *included* variables are, tautologically, captured by those variables. In the case of Proposal Six, TPF (in natural logs, along with the first and twelfth lags) are included in the observed variables x_{it} and the coefficients on those variables capture all the effects of TPF on workhours.

The cited statement from Docket No. R97-1 thus misinterprets the one-way fixed-effects model with site (individual) effects. In a one-way fixed-effects model, the site-specific dummy variables (or equivalently site-specific intercepts) are constants over time by construction, and thus cannot be affected by variations in volume. As noted above, they estimate the effects on workhours of otherwise omitted site-specific variables. Indeed, the Commission correctly observed in Docket No. R97-1 that the site-specific intercepts or dummy variables capture the effects of factors that are invariant over the regression sample period (cf. PRC Op., R97–1, Vol. 1, at 86; Vol. 2, Appendix F, at 10). A "volume-variable indirect effect" of TPF captured by the fixed-effects terms must be an effect varying on the margin with respect to TPF (otherwise it is not volume-variable) while being invariant over the sample period (otherwise it is not captured by the fixed-effects terms). Any such indirect effect must also not be captured by the (log) TPF variable itself, or its lags, which are included variables in the Proposal Six models. This combination of features requiring the effects to be simultaneously constant over time and variable with respect to TPF represents a logical contradiction and mathematical impossibility.

RESPONSE OF THE UNITED STATES POSTAL SERVICE TO CHAIRMAN'S INFORMATION REQUEST NO. 6

6. Please refer to the Variability Report that states "[t]he FY2016-FY2019 period features a relatively fixed operating environment including technology mix, while providing sufficient regression sample sizes, and serves as the sample period for the main estimation results." Variability Report at 21. Please also refer to the Response to CHIR No. 3 that states "[l]imiting the amount of time variation in factors such as management quality, facility layouts, or local demographics is a partial motivation for employing a relatively short time period for the regression sample periods—i.e., the proposed four-year period rather than the full FY2007-FY2019 period." Response to CHIR No. 3, question 3.b.

a. Please confirm that using a shorter time period such as a 3-year period would further limit the amount of time variation in factors such as management quality, facility layouts, or local demographics and improve the ability of the fixed effects estimator to account for unobserved non-volume heterogeneity among facilities.

b. If question 6.a. is not confirmed, please explain the effect of using a shorter time period on the applicability of a fixed effects estimator.

RESPONSE:

a. Confirmed. However, limiting the amount of time variation in unobserved non-volume factors affecting workhours is not the only consideration in determining the period of analysis. Other factors include providing adequate sample size, currency of the regression data, and the stability of results over consecutive sample periods. The four-year period in Proposal Six was intended to balance those considerations. Shortening the sample period would make the regression sample more current, while reducing sample sizes and potentially leading to less stable elasticities over time. A shorter time period such as three years may represent a reasonable alternative to Proposal Six — and a superior alternative to the existing variability assumptions — if the Commission were to weigh factors favoring a shorter time period more heavily.

b. Not applicable.

7. Please refer to the Variability Report that states "[t]he FY2016-FY2019 period features a relatively fixed operating environment including technology mix, while providing sufficient regression sample sizes, and serves as the sample period for the main estimation results." Variability Report at 21. Please also refer to Table 2 below that provides two sets of variabilities: (1) those estimated in Proposal Six over the FY 2016-2019 sample period[3] and (2) variabilities derived from the workhour regression equations estimated over a FY 2017-2019 sample period. All other assumptions underlying Proposal Six, including a 5% productivity screen, were kept the same.

a. Please confirm that variability estimates provided in column (1) of Table 2 are estimated correctly. If not confirmed, please provide the corrected variabilities and explain the reasons for the occurred discrepancies. With your response please include program, log, and output files.

b. If question 7.a. is not confirmed and, if the corrected variabilities estimated in question 7.a. are different from variabilities estimated in Proposal Six (see column (3) of Table 2), please explain the reasons why the variabilities changed when they were estimated over a slightly shorter time period (considering that the FY 2016-2019 sample period featured a relatively fixed operating environment as suggested in the Variability Report). Variability Report at 21.

c. If question 7.a. is confirmed, please explain why variabilities changed quite substantially when estimated over a slightly shorter time period (considering that the FY 2016-2019 sample period featured a relatively fixed operating environment as suggested in the Variability Report). *Id.*

Table 2. Workhour Variability Estimates for Alternative Sample Periods

	5% screen, FY2017-2019		5% screen, FY2016-2019 (Proposal Six)	
	coeff	se	coeff	se
	(1)	(2)	(3)	(4)
DBCS	0.925	0.025	0.976	0.032
AFSM 100	0.850	0.085	0.774	0.091
FSS	0.789	0.080	0.804	0.070

Notes and Sources: Data for FY 2017-2019 sample period estimates (columns (1) and (2)) are from Library Reference USPS-RM2020-13/1, folder "Analysis," data file "analysis_set.dta." The sample period was modified by substituting the STATA code: "inrange(year, 2017, 2019)" for "inrange(year, 2017, 2019)" in STATA do file "analysis_seasonal.do" located in Library Reference USPS-RM2020-13/1, folder

[3] *See* Library Reference USPS-RM2020-13/1, folder "Analysis," Excel file "results_seasonal.xlsx."

"Analysis." Proposal Six estimates (columns (3) and (4)) are from Library Reference USPS-RM2020-13/1, folder "Analysis," Excel file "results_seasonal.xlsx."

RESPONSE:

a. Confirmed.

b. Not applicable.

c. While the question characterizes the 2017-2019 time period as "slightly shorter" than the 2016-2019 period used in Proposal Six, the effect of dropping one year from the period is to reduce the sample sizes for the regressions by approximately 25 percent for each operation. Notwithstanding the relatively large change in sample size, the elasticity estimates based on 2017-2019 data are all within the 95 percent confidence intervals of the Proposal Six estimates. The 2017-2019 estimates for the AFSM 100 and FSS operations are within one standard error of the Proposal Six estimates, and the elasticity difference of 0.015 for FSS is not large. The 2017-2019 results also are consistent with the qualitative feature of Proposal Six that the DBCS elasticity remains materially closer to 100 percent volume-variability than the AFSM 100 and FSS elasticities.

   As noted in the response to Question 6(a) of this Information Request, a potential consequence of shortening the sample period is to increase the volatility of the estimates in successive sample periods. In the Variability Report at p. 26, the rolling-sample analysis showed that elasticities from successive four- and five-year (48- and 60-month) sample periods exhibited broadly similar trends, though the paths of elasticities from the five-year samples are somewhat smoother than those of the four-year samples. Figure 1, below, shows comparable results from three- and four-year rolling sample periods. An Excel file with the figure, and the Stata code and output log generating the data, are electronically attached to this response. As with the four- and five-year results in the Variability Report, the elasticities for each group generally track each other over time, though the elasticities from the three-year periods tend to have earlier and more pronounced fluctuations than the four-year period.

RESPONSE OF THE UNITED STATES POSTAL SERVICE TO CHAIRMAN'S INFORMATION REQUEST NO. 6

Figure 1.